International Journal of Computer Vision, 4, 39–57 (1990)

Detecting Moving Objects*

WILLIAM B. THOMPSON AND TING-CHUEN PONG
Computer Science Department, University of Minnesota, Minneapolis, MN 55455

Abstract
The detection of moving objects is important in many tasks. This paper examines moving object detection based primarily on optical flow. We conclude that in realistic situations, detection using visual information alone is quite difficult, particularly when the camera may also be moving. The availability of additional information about camera motion and/or scene structure greatly simplifies the problem. Two general classes of techniques are examined. The first is based upon the motion epipolar constraint—translational motion produces a flow field radially expanding from a "focus of expansion" (FOE). Epipolar methods depend on knowing at least partial information about camera translation and/or rotation. The second class of methods is based on comparison of observed optical flow with other information about depth, for example from stereo vision. Examples of several of these techniques are presented.

1 Introduction

One important function of a vision system is to recognize the presence of moving objects in a scene. If the camera is stationary and illumination constant, this can be done by simple techniques that compare successive image frames looking for significant differences. If the camera is moving, the problem is considerably more complex. For the purposes of this discussion, *moving objects* are taken to be any objects moving with respect to the stationary portions of the scene, which we refer to as the *environment*. For a moving camera, both moving objects and stationary portions of the scene may be changing position with respect to the camera and thus generating visual motion in the imagery. A moving camera leads to difficulties because of the need to determine objects moving with respect to the environment, rather than the much easier problem of finding objects moving with respect to the camera. In this article, we deal with the problem of detecting moving objects from a moving camera based on optical flow.

The visual detection of moving objects is a surprisingly difficult task. A simple example illustrates just how serious the problem can be. Consider the optical flow field shown in figure 1, which appears to show a small, square region in the center of the image moving to the right and surrounded by an apparently stationary background. Such a flow field can arise from several equally plausible situations: (1) The camera is stationary with respect to the environment, and the central region corresponds to an object moving to the right. (2) The camera is moving to the left with respect to the environment, most of the environment is sufficiently distant so that the generated optical flow is effectively zero, while the central region corresponds to a surface near to the camera but stationary with respect to the environment. (3) The camera and object are moving with respect to both the environment and each other, though the environment is sufficiently distant so that there is

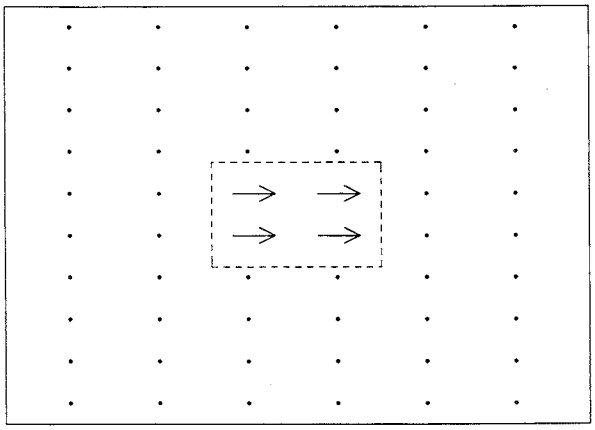

Fig. 1. Is the central region a moving object?

*A preliminary version of this article appeared in *The Proceedings of the First International Conference on Computer Vision*, London, June 1987.

no perceived optical flow. It is not possible to tell whether or not this seemingly simple pattern corresponds to a moving object![1]

Figure 1 provides one example of why a general and reliable solution to the problem of moving object detection based only on optical flow is not feasible. Robust solutions require that additional information about camera motion and/or scene structure be available. In this article, we examine three types of information that might be available. Each information source places constraints on the optical flow fields that can be generated by a camera moving through an otherwise static environment. Violations of these constraints are thus necessarily due to moving objects. One group of techniques depends on knowing partial information about how the camera is moving. A second approach requires that an object of interest be actively tracked by the camera system. The final group of methods is based on knowing information about the depth to surface points, either directly from approaches such as stereo or indirectly via a priori knowledge that object motion takes place over smooth surfaces.

At least three general approaches to moving object detection are possible. Each exploits a particular constraint that must hold if a camera is moving through an otherwise static environment. Detecting moving objects becomes equivalent to a search for violated constraints.

- *Motion epipolar constraint:* Translational camera motion produces a distinctive optical flow pattern. Flow vectors appear to radiate out from a "focus of expansion" (FOE) corresponding to the line of sight coincident with the direction of motion. This has the effect of constraining the orientation of flow vectors. Visual motion that violates this orientational constraint must be due to moving objects. Under some circumstances, the motion epipolar constraint may still be used when camera rotation is added to the translational movement.
- *Depth/flow constraint:* The optical flow generated by a surface point is a function of the relative motion between camera and surface and of the range to the surface. If range values are available, then inconsistencies between optical flow, range, and observer motion signal moving objects.
- *Rigidity constraint:* A scene containing moving objects can be thought of as undergoing nonrigid motion with respect to the camera. Structure-from-motion techniques which are sensitive to the presence of nonrigid motion can thus be used to detect moving objects.

[1]The flow pattern in figure 1 provides little information about actual camera motion. Apparently stationary image regions can be due to the viewing of distant surfaces and/or rotational motion that tracks a surface point, keeping it at a fixed point in the field of view. Even with significant nonzero flow existing over the whole of the image, ambiguities exist between flow patterns due to translational motion and due to rotational motion [1].

This paper will concentrate on epipolar and depth/flow methods. Though potentially effective, methods based directly on the rigidity constraint require longer frame sequences, temporal derivatives of optical flow, and/or a wide field of view to enhance perspective effects.

Many theoretically plausible techniques for analyzing visual motion are ineffective in practice. Typically, the assumptions on which these techniques are either explicitly or implicitly founded do not accurately represent real problems. For this work, we start with the presumption that motion detection algorithms should be designed with the following properties in mind:

- *The field of view may be relatively narrow:* Motion detection should not depend on the use of wide-angle imaging systems. Such systems may not be available in a particular situation, and if used may increase the difficulty of recognizing small moving objects. As a result, detection algorithms should not depend on subtle properties of perspective.
- *The image of moving objects may be small with respect to the field of view:* This is clearly desirable for reliability. Moving objects may be far away and subtended by relatively small visual angles. We need methods capable of identifying single image points, or at least small collections of points, as corresponding to moving objects. Detection algorithms thus cannot depend on variations in flow over a potentially moving object.
- *Estimated optical flow fields will be noisy:* No method is capable of estimating optical flow with arbitrary accuracy. Motion detection based on optical flow must be tolerant of noisy input.

2 Background

An extensive literature has developed on computational approaches to the analysis of visual motion (e.g., see [2]). The majority of this work deals with what Ullman

[3] has called the structure-from-motion and motion-from-structure problems. Visual motion is used to determine the three-dimensional position of surface points under view and/or the parameters of motion relating camera and object. Almost without exception, papers describing structure-from-motion and motion-from-structure algorithms deal only with a single, rigid object in the field of view. If the problem of separately moving objects is mentioned at all, it is in a comment that the image must be segmented into separately moving objects before the method being described is applied.

Some work has been done on the segmentation of images based on visual motion. The easiest form of this problem occurs with a camera known to be stationary. In such circumstances, object motion leads to significant temporal differences in an image sequence. Such differences correspond to moving objects, and furthermore can be used to estimate the boundaries of the objects (e.g., [4, 5]). More classical edge-detection techniques can also be applied to time-varying imagery [6, 7, 8, 9, 10, 11]. Such approaches work for both moving and stationary cameras. When the camera is moving, however, sharp spatial changes in visual motion can correspond to either the boundaries of moving objects or to depth discontinuities between two rigidly attached surfaces. As a result, motion-based edge detection is not sufficient to detect moving objects. In particular, moving-object detection in the general case is a very different and more difficult problem than motion-based segmentation.

Ullman suggested the use of a rigidity check to segregate a scene into features corresponding to separately moving objects [3]. Jain explicitly dealt with the problem of detecting moving objects using a moving camera [8]. His approach exploited the motion epipolar constraint which says that for translational camera motion with respect to a static environment, optical flow will expand radially from a focus of expansion corresponding to the direction of translation. For translational motion, any flow values violating the epipolar constraint must be due to moving objects in the scene. Unfortunately, this approach requires knowledge of the direction of translation and does not work if the motion has a rotational component. Heeger and Hager [12] and Zhang et al. [13] find moving objects by solving for the parameters of camera motion and then finding areas of the scene that move relative to the camera in a manner inconsistent with these estimated parameters. This approach may have difficulty in situations such as narrow viewing angles in which accurate recovery of camera motion is difficult. In addition, Zhang et al. require a calibrated stereo system.

3 The Optical Flow Equation

The basic mathematics governing the optical flow generated by a moving camera is well known. Our notation is similar to [14], using a coordinate system fixed to the camera (e.g., the world can be thought of as moving by a stationary camera). Optical flow values are a function of image location, the relative motion between the camera and the surface point corresponding to the image location, and the distance from the camera to the corresponding surface point:

$$\mathbf{F}(\mathbf{p}) = \frac{\mathbf{F}_t(\mathbf{p})}{r(\mathbf{p})} + \mathbf{F}_r(\mathbf{p}) \tag{1}$$

$$\mathbf{F}_t = (-U + xW,\ -V + yW) \tag{2}$$

$$\mathbf{F}_r = [Axy - B(x^2 + 1) + Cy,\ A(y^2 + 1) - Bxy - Cx)] \tag{3}$$

where $\mathbf{F}$ is the optical flow at image location $\mathbf{p} = (x, y)$, x and y are normalized by the focal length, $r(\mathbf{p})$ is the range from the camera to the surface point imaged at $\mathbf{p}$, $\mathbf{T} = (U, V, W)^T$ specifies the translational velocity of the camera, and $\omega = (A, B, C)^T$ specifies camera rotation.

Most work on the analysis of optical flow has dealt with a camera moving through an otherwise static environment or, equivalently, a single rigid object moving in front of a fixed camera. In such cases, single values of $\mathbf{T}$ and ω govern the flow over the whole image. If moving objects are present, then the relative motion between camera and environment will be different than the relative motion between camera and moving object. Notationally, we will specify the camera motion with respect to the environment by $\mathbf{T}^{(\text{env})}$ and $\omega^{(\text{env})}$. The parameters specifying the relative motion between the camera and an arbitrary scene point $\mathbf{p}$ will be indicated by $\mathbf{T}^{(\mathbf{p})}$ and $\omega^{(\mathbf{p})}$. $\mathbf{p}$ lies on a moving object if $\mathbf{T}^{(\mathbf{p})} \neq \mathbf{T}^{(\text{env})}$ and/or $\omega^{(\mathbf{p})} \neq \omega^{(\text{env})}$.

4 Detection Based on Epipolar Constraint

If complete information about instantaneous camera motion is available, then $\mathbf{T}^{(\text{env})}$ and $\omega^{(\text{env})}$ are known. If the camera is translating but not rotating with respect

to the background, $\omega^{(\text{env})} = 0$, $\mathbf{F}_r = 0$, and all flow vectors due to the moving image of the background will radiate away from a *focus of expansion* (FOE). From equations (1) and (2) it is easy to see that the image plane location of the FOE is at

$$(x, y)_{\text{FOE}} = \left(\frac{U}{W}, \frac{V}{W}\right) \tag{4}$$

While the location of the FOE depends only on the direction of translation and not on the speed, it is important for detectability that the speed be sufficient to generate measurable optical flow. The FOE is not restricted to lie within the visible portion of the image (and in fact may be a focus of contraction). An FOE at ∞ corresponds to pure lateral motion, which generates a parallel optical flow pattern.

4.1 Direct Use of Motion Epipolar Constraint for Known Camera Motion

For pure translational motion, the direction of motion specifies the direction of optical flow associated with any surface point stationary with respect to the environment:

$$\theta_{\text{FOE}} = \tan^{-1}\frac{V - Wy}{U - Wx} \tag{5}$$

where θ_{FOE} is the expected flow orientation at the point (x, y), predicted using the motion epipolar constraint. Note that this equation is still well defined when $W = 0$, corresponding to a focus of expansion at ∞ in image coordinates. Any flow values with a significantly different direction correspond to moving objects [8]. (The converse is not necessarily true. It is possible that moving objects coincidentally generate flow values compatible with this constraint.) This approach requires the estimation of only the direction of flow, not either the magnitude or spatial variation of flow.

Camera rotation introduces considerable complexity. Knowledge of camera motion no longer constrains the direction of background flow. Nevertheless, at a given point $\mathbf{p}$, flow is constrained to a one-dimensional family of possible vector values. The family is given by equations (1)–(3) where r ranges over all positive values. The analysis can be simplified because of the linear nature of (1). $\mathbf{F}_r$ depends only on the parameters of rotation and not on any shape property of the environment. Because the value of $\mathbf{F}_r$ at a particular $\mathbf{p}$ does not depend on $r(\mathbf{p})$, it can be predicted knowing only ω. At every point within the field of view, this value can be subtracted from the observed optical flow field, leaving a *translational flow field*:

$$\mathbf{F}_{\text{trans}} = \mathbf{F} - \mathbf{F}_r \tag{6}$$

This field behaves just as if no rotation was occurring, and thus moving objects can be located using the FOE technique described above. For the remainder of this paper, when rotation is present, we will take the term FOE to refer to the focus of expansion of this translational field.

In principle, even if camera motion is not known $\mathbf{T}^{(\text{env})}$ and $\omega^{(\text{env})}$ may be estimated from the imagery (e.g., [14]), subject to a positive, multiplicative scale factor for $\mathbf{T}^{(\text{env})}$. Two serious problems exist, however. Narrow angles of view make estimation of camera motion difficult, as significantly different parameters of motion and surface shape can yield nearly identical optical flow patterns [1]. In addition, techniques such as [14] use a global minimization approach which will not perform well if moving objects make up a substantial portion of the field of view. A clustering approach (e.g., [15]) can be made tolerant of the moving objects, though great difficulty can be expected dealing with a five-dimensional cluster space.

4.2 Indirect Use of Motion Epipolar Constraint

The motion epipolar constraint has an important implication for motion analysis methods that operate only over small image neighborhoods. Away from the FOE, $\mathbf{F}_t(\mathbf{p})$ and $\mathbf{F}_r(\mathbf{p})$ vary slowly with $\mathbf{p}$ (see equations (2) and (3)). Over a small neighborhood, both $\mathbf{F}_t(\mathbf{p})$ and $\mathbf{F}_r(\mathbf{p})$ are essentially constant. As a result, over a small neighborhood, the component of flow due to rotational motion is essentially constant, while the translational flow, $\mathbf{F}_{\text{trans}}$, varies only by a scalar multiple dependent on depth. That is, over the neighborhood $\mathbf{F}_{\text{trans}}$ is essentially constant in direction. We can use this result to simplify problems arising from rotational camera motion. In one technique, we explicitly compensate for rotation. In a second technique, active tracking of potentially moving objects leads to a particularly simple computational scheme.

4.2.1 Known Rotation. Often, information about camera rotation is available, even when the direction of translation is not known. Nonvisual information about camera motion often comes from inertial sources. Such sources

are much more accurate in determining rotation than translation. Rotation involves a continuous acceleration which is easily measured. The determination of translation requires the integration of accelerations, along with a starting boundary value. Errors in estimated translation values rapidly accumulate. A simple technique allows the detection of moving objects when only camera rotation is known.

If all motion parameters are known, knowledge of camera rotation makes it possible to compute the translational flow field, $\mathbf{F}_{\text{trans}}$. Knowledge of translation can then be used to locate the FOE and thus constrain the direction-of-flow vectors associated with the environment. If only rotation is known, it is still possible to determine the translational flow field, but not the FOE. Visual methods can be applied to the translational flow field to estimate the location of the FOE, but these methods suffer from a number of practical limitations when applied to noisy data.

An alternate approach can be used which does not require the prior determination of the FOE. The translational flow field extends radially from the focus of expansion. From the arguments given above, we know that over any local area away from the FOE, variations in the *direction* (but not necessarily magnitude) of the translational flow field will be small. Flow arising due to moving objects is of course not subject to this restriction. The gradient of flow field direction can thus be used to detect the boundaries of moving objects. At these boundaries, flow direction will vary discontinuously.[2]

A complementary technique is available to deal with situations in which translation but not rotation is known. We can expect these situations to be rare, however. If the direction of translation were known over some interval of time, it would be an easy matter to determine the rotation by examining the rate of change of direction.

4.2.2 Active Tracking. A vision system that can actively control camera direction is capable of tracking regions of interest over time, keeping some particular object centered within the field of view. Tracking regions of interest is desirable for many reasons other than the detection of moving objects (e.g., [17]), though the analysis of imagery arising from a tracking camera has not received much study by the computer vision community. If there are significant variations in depth over the visible portion of the background and if moving objects are relatively small with respect to the field of view, then moving object detection based on tracking can be accomplished without any actual knowledge of camera motion. (For motion detection, the tracking can easily be simulated if the camera is not actively controllable.)

If an object is being tracked, then its optical flow is zero.[3] Flow-based methods for determining whether or not a tracked object is moving must depend wholly on the patterns of flow in the background. Object tracking helps in moving object detection because it minimizes many of the difficulties due to camera rotation. When dealing with instantaneous flow fields, we can decompose the problem by considering all translational motion to be due to movement of the camera platform and all rotational motion due to pan and tilt of the camera to accomplish the tracking. (We will disregard any effects due to spin around the line of sight.) Consider the effect of tracking a point that is in fact part of the environment. Tracking is effected by generating a rotational motion that exactly compensates for the translational flow at the center of the image. This is accomplished by choosing $\mathbf{F}_r$ such that:

$$\mathbf{F}_r(0, 0) = -\frac{\mathbf{F}_t(0, 0)}{r(0, 0)} \tag{7}$$

For a small enough neighborhood, $\mathbf{F}_t$ and $\mathbf{F}_r$ can be treated as constant, leading to the following flow equation:

$$\mathbf{F}_{\text{track}}(\mathbf{p}) = \left(\frac{1}{r(\mathbf{p})} - \frac{1}{r(0, 0)}\right)\mathbf{F}_t \tag{8}$$

The effect on the optical flow field is that in the neighborhood of the tracked point, the direction of flow will be approximately constant (modulo 180°), with a magnitude dependent on the difference between the range to the corresponding surface point and the range to the tracked point.

Now, consider tracking a point that is moving with respect to the environment. If environmental surface points are visible in the neighborhood of the tracked point, $\mathbf{F}_t$ and $\mathbf{F}_r$ are no longer constant within the neighborhood. For environmental points:

[2]Marr [16] claims "if direction of [visual] motion is ever discontinuous at more than one point—along a line, for example—then an object boundary is present." Note that this is only necessarily true if no camera rotation is occurring (or equivalently, if camera rotation has been normalized by using the translational flow field).

[3]To simplify discussion, we ignore the case of an object rotating at constant depth. The method developed does in fact deal effectively with this situation.

$$\mathbf{F}_{\text{track}}(\mathbf{p}) = \frac{\mathbf{F}_t^{(\text{env})}}{r(\mathbf{p})} + \mathbf{F}_r^{(\text{env})} - \frac{\mathbf{F}_t^{(\text{object})}}{r(0, 0)} \quad (9)$$

$\mathbf{F}_t^{(\text{env})}$, $\mathbf{F}_r^{(\text{env})}$, and $\mathbf{F}_t^{(\text{object})}$ will in general differ in orientation. If there is a variation in range to visible environmental points, then there will be a variation in direction of observed flow over the neighborhood. (Note that detection is not possible if there is no variation in $r(\mathbf{p})$ over the visible environment. This situation is similar to that depicted in figure 1.)

Figures 2 and 3 illustrate the effect. Figure 2 shows the optical flow over a neighborhood in which no motion is occurring with respect to the environment. Figure 2a shows the flow before any tracking motions are initiated. The dashed line indicates the translational component of flow. The rotational component of flow is indicated by the dotted line. The solid line is the observed optical flow, the sum of the translational and rotational components. The translational components are parallel. The variations in magnitude correspond to underlying variations in range. The rotational components are constant over the neighborhood. Note that the observed flow varies in orientation—as previously indicated, orientational variability alone is not enough to detect moving objects. Figure 2b shows the flow that results when the point in the center of the region is being tracked.

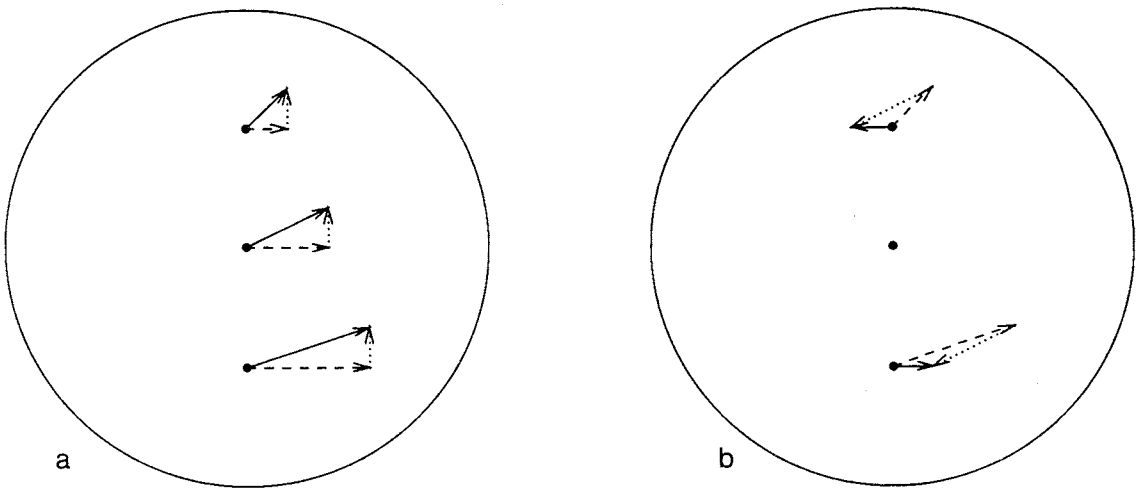

Fig. 2. Tracking a stationary surface point. (a) Before tracking. (b) After tracking.

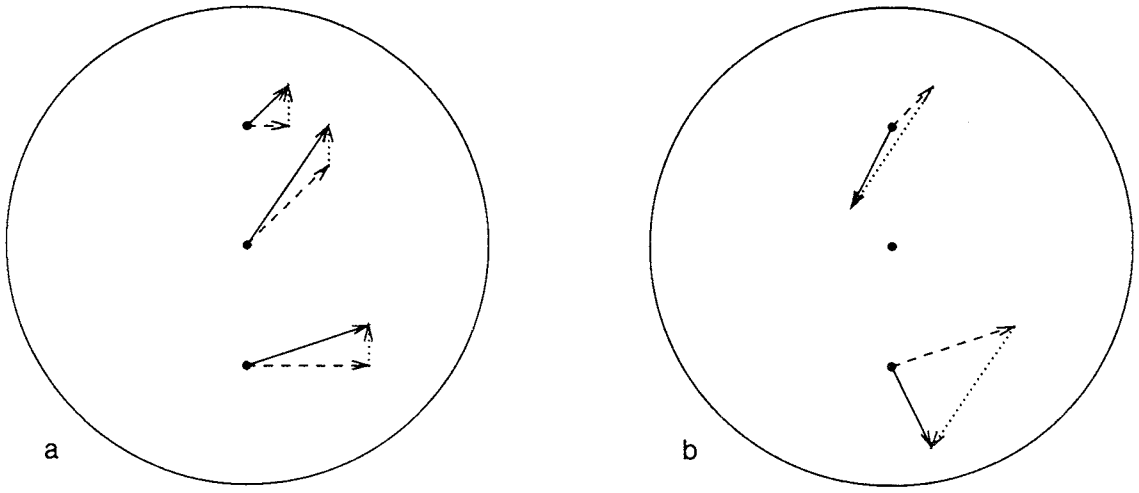

Fig. 3. Tracking a moving object. (a) Before tracking. (b) After tracking.

The center flow is of course zero. The dashed lines now indicate the flow that would be observed without tracking. The dotted lines indicate the rotational flow that is introduced to stabilize the center point within the field of view. The solid line shows the resulting optical flow. Note that the flow vectors are parallel, but in this case differ by 180°.

Figure 3 shows the same flow vectors in the case where the center point corresponds to a moving object and the two other points correspond to portions of the environment. Note that in figure 3a, the translational flow varies significantly in orientation. If we actually knew the translational flow, this fact would be enough to determine that a moving object was present. Without information about camera rotation, however, we must resort to more indirect methods.

5 Detection Based on Flow/Depth Constraint

Recently, efforts have been made to develop integrated approaches to analyzing stereo and motion (e.g., [6, 18]). These approaches simultaneously deal with motion and stereo disparity, either by comparing flow fields taken from different viewing positions or by establishing point correspondences over both time and viewing directions. Similar multi-cue analysis can greatly aid in the detection of moving objects. We claim, however, that it is not necessary to adopt a strategy requiring the unified low-level integration of motion and stereo. Rather, depth estimates from whatever sources are available and can be used. In addition to stereo, these sources can include the full range of nonmotion depth cues: familiar size, focus, gradients of various properties, aerial perspective, and many more [19]. Furthermore, while precise estimates of depth are obviously useful, relative depth or coarse approximations to depth can also aid in the analysis.

5.1 Objects Moving on Surfaces

Knowledge of the shape of environmental surfaces can be used to simplify the motion detection problem. Scene structure may be known precisely (e.g., the range to visible surface points) or in terms of general properties (e.g., significant depth discontinuities can be expected). If moving objects must remain in contact with environmental surfaces (e.g., vehicular motion), a less complex technique is possible; it depends only on knowing the image plane locations corresponding to discontinuities in range. If no objects are moving within the field of view, equations (1)–(3) show that flow varies inversely with distance for fixed $\mathbf{p}$. Both $\mathbf{F}_r$ and $\mathbf{F}_t$ vary slowly (and continuously) with $\mathbf{p}$. Discontinuities in $\mathbf{F}$ thus correspond to discontinuities in r. This relationship holds only for relative motion between the camera and a single, rigid structure. When multiple moving objects are present, equation (1) must be modified so that there is a separate $\mathbf{F}_r^{(i)}$ and $\mathbf{F}_t^{(i)}$ specifying the relative motion between the sensor and each rigid object. Discontinuities in flow can now arise due either to a discontinuity in range or to the boundaries of a moving object. If independent information is available on the location of range discontinuities, other discontinuities in flow must be due to moving objects.

The motion detection problem becomes particularly simple if the environment is planar. In this case, depth discontinuities are not possible and *any* discontinuity in flow (either direction or magnitude) corresponds to the boundary of a moving object. Note that it is not sufficient to know simply that the environment is a "smooth" surface. From some viewing positions, even smooth surfaces may exhibit range discontinuities.

5.2 Direct Comparison of Depth and Flow

A simple way of combining depth and visual motion to detect moving objects is possible if accurate 3-D position information is available for a sufficient number of surface points in the environment and on any moving objects. If both the optical flow and the depth are known for a collection of surface points in the environment, then equations (1)–(3) can be used to create a system of equations which can be solved for the parameters of motion $\mathbf{T}^{(\text{env})}$ and $\omega^{(\text{env})}$. (Knowing depth values makes this an easier task than the standard structure-from-motion problem.) If the collection of points includes some values associated with the environment and others associated with one or more objects moving with respect to the environment, the system of equations used to solve for $\mathbf{T}$ and ω will be inconsistent. Checking the system for consistency can therefore be used as a test for the presence of a moving object (e.g., a test for nonrigid motion in the field of view). Only the consistency of the system is important. The actual values of $\mathbf{T}$ and ω are not relevant to the detection problem.

5.3 Indirect Comparison of Depth and Flow

The availability of accurate 3-D position estimates depends in large part on having accurately calibrated camera systems. Not only is this calibration difficult, but it is continuously subject to variability due to mechanical compliance. *Relative* measures of visible motion and/or stereo can be used to avoid this calibration problem (e.g., [20]). For example, Reiger and Lawton have shown how to use local spatial differences to minimize difficulties due to rotation [21]. If no moving objects are visible, then large local differences in flow can only be due to a change in depth. If $\mathbf{p}^{(i)}$ and $\mathbf{p}^{(j)}$ are image points on either side of such a boundary, then from equation (1) we have

$$\Delta\mathbf{F} = \left\| \mathbf{F}_r(\mathbf{p}^{(i)}) - \mathbf{F}_r(\mathbf{p}^{(j)}) + \frac{\mathbf{F}_t(\mathbf{p}^{(i)})}{r(\mathbf{p}^{(i)})} - \frac{\mathbf{F}_t(\mathbf{p}^{(j)})}{r(\mathbf{p}^{(j)})} \right\| \tag{10}$$

If $\mathbf{p}^{(i)}$ and $\mathbf{p}^{(j)}$ are sufficiently close, $\mathbf{F}_r(\mathbf{p}^{(i)}) \approx \mathbf{F}_r(\mathbf{p}^{(j)})$ and $\mathbf{F}_t(\mathbf{p}^{(i)}) \approx \mathbf{F}_t(\mathbf{p}^{(j)})$. As a result the rotational component of flow cancels out in the spatial difference and

$$\Delta\mathbf{F} \approx \left\| \mathbf{F}_t(\mathbf{p})\, \Delta\left(\frac{1}{r}\right) \right\| \tag{11}$$

That is, the difference in flow across the edge is proportional to the difference of the reciprocal of depth across the edge. The relationship between stereo disparity and depth is very similar to the relationship between optical flow and depth:

$$d(\mathbf{p}) = d_v(\mathbf{p}) + \frac{d_b(\mathbf{p})}{r(\mathbf{p})} \tag{12}$$

where $d(\mathbf{p})$ is the stereo disparity at $\mathbf{p}$, d_v is a term dependent on the camera vergence, and d_b is a term dependent on the baseline separating the cameras. Using the same argument as above, we have

$$\Delta d \approx \left\| d_b(\mathbf{p})\, \Delta\left(\frac{1}{r}\right) \right\| \tag{13}$$

Over a local neighborhood, $\mathbf{F}_t$ and d_b will remain essentially constant, while $\Delta(1/r)$ will generally vary. Dividing equation (11) by equation (13) shows that the *ratio* of $\Delta\mathbf{F}$ to Δd remains constant, as long as the points over which the differences are taken are the same for flow and disparity.

Flow boundaries associated with moving objects are not subject to this constraint. As a result, we can detect moving objects by looking for local neighborhoods over which the ratio $\Delta\mathbf{F}/\Delta d$ varies significantly. We never have to solve for the actual depth, nor do we need to know the functions $\mathbf{F}_t$, $\mathbf{F}_r$, d_v, or d_b. The solution does not depend on information about camera motion or relative camera geometry. For this approach to work, however, there must be significant changes in depth over the background, not just between the background and any moving objects. There is reason to believe that such variation is important to a large class of moving object detection algorithms.

6 Examples

All of the methods described in Sections 4 and 5 have been tested experimentally. Four examples are presented in this section, all involve a moving camera and potentially moving objects. Two cases exploit the epipolar constraint. The first of these involves a situation in which camera rotation is known, but not camera translation. In the second case, a potentially moving object is being actively tracked. Results are also presented for two methods utilizing constraints resulting from the comparison of depth and flow. The simplest of these involves objects moving over a smooth environment. The final example compares flow and disparity across boundaries of possibly moving objects, using the technique of Section 5.3.

Figure 4 shows the first frame in a sequence of images of an outdoor scene. In this example, the camera rotates and translates with respect to the environment while the toy vehicle moves to the right between image frames. The rotational velocity of the camera with respect to the environment was measured. The optical flow field shown in figure 5 was obtained by the token matching technique described in [22]. The translational flow field shown in figure 6 was obtained by subtracting the rotational flow component computed from the known rotational velocity from the observed optical flow field (figure 5). The gradient of flow direction in the translational flow field was used to detect the boundaries of moving objects. Figure 7 shows the detected boundary of a moving object overlaid onto the first frame of figure 4.

In figure 8 the mechanical toy creature in the center of the image is being tracked by the camera while the camera is translating to the left with respect to the environment. Figure 9 shows the estimated optical flow. Figure 10 shows a histogram of the directions of the optical flow. Note that there are two distinct peaks in the histogram. The variation in flow direction over the

Fig. 4. First frame of outdoor sequence.

Fig. 5. Optical flow field obtained from the image sequence of figure 4.

Fig. 6. Translational flow field determined from the optical flow field of figure 5.

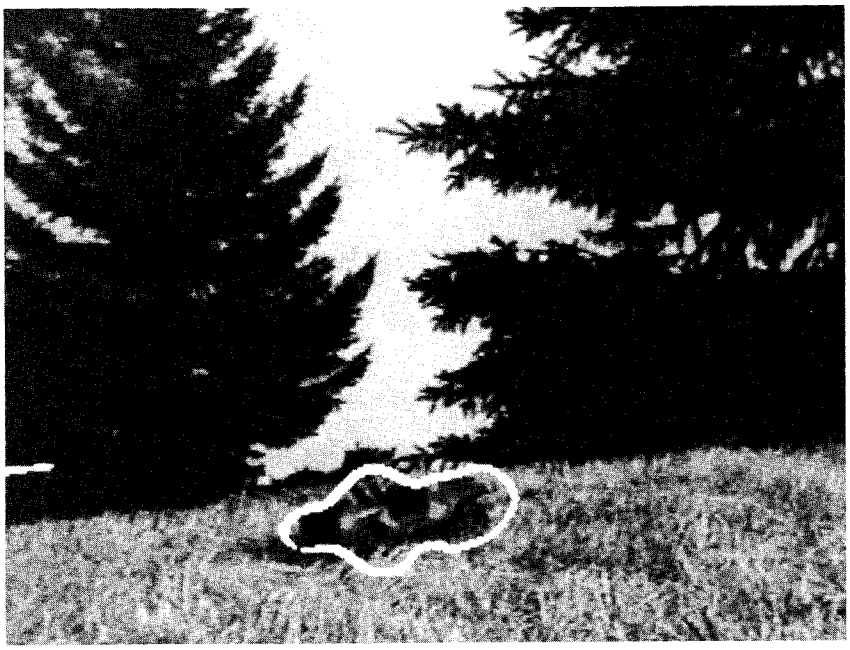

Fig. 7. Boundary of a moving object overlaid onto the first image of figure 4.

Fig. 8. First frame of tracking sequence.

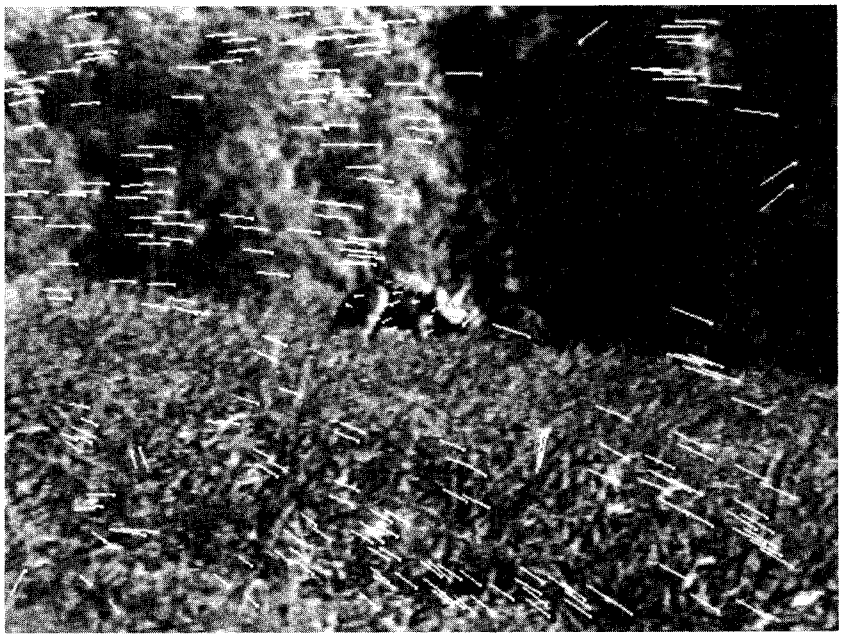

Fig. 9. Optical flow field obtained from the image sequence of figure 8.

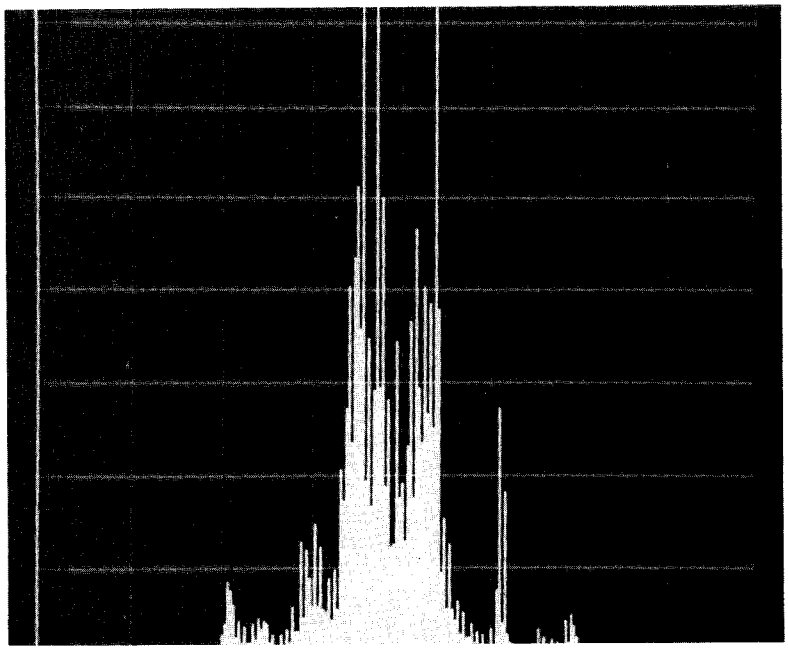

Fig. 10. Histogram of the flow direction of the optical flow vectors in figure 9.

image was computed to be approximately 34°, indicating that the tracked object was in fact moving.

As a comparison, a similar experiment in which the tracked object, a rock, is stationary with respect to the environment while the camera is moving was also performed. A pair of images similar to that of figure 8 were obtained. The resulting estimated optical flow field is shown in figure 11. Its corresponding histogram is shown in figure 12. Note that only one distinct peak is observed in this histogram. The global variation in flow direction in this case was computed to be approximately 11° which is significantly smaller than that of the previous example.

An image sequence starting with the frame shown in figure 13 is used to illustrate the technique for detecting objects moving in a smooth environment. In this example, the camera moves with respect to an environment consisting of various small pieces of hardware lying on a planar surface. The optical flow field shown in figure 14 was obtained in the same manner as in figure 5. Figure 15 shows the locations of large variations in optical flow values, corresponding to the boundary of a moving object.

A stereo image sequence starting with the stereo pair shown in figure 16 is used to illustrate the technique of indirect comparison of flow and disparity as a basis for moving object detection. Both the flow field shown in figure 17, and the disparity field shown in figure 18 were obtained using the method of figure 5. Comparing the ratio of the change in disparity values to the change in flow values across neighboring points, and selecting as the boundaries of moving objects those areas in which there is a distinct discontinuity in that ratio, results in the identification of the boundaries indicated in figure 19.

7 Discussion

7.1 Which Method to Use?

This paper presents a collection of loosely related techniques for visually detecting moving objects. Detection based purely on visual motion from a single camera seems quite difficult. Each of the methods presented here uses some sort of additional information, about either current camera motion or scene structure. The methods are characterized by the additional information used, the underlying constraints exploited, and the particular computational structure used to implement the technique. It is likely that reliable moving object detection will require several complementary techniques, along with a method for selecting which detector to trust in any particular situation.

Fig. 11. Optical flow field obtained from tracking an object that is stationary with respect to the environment.

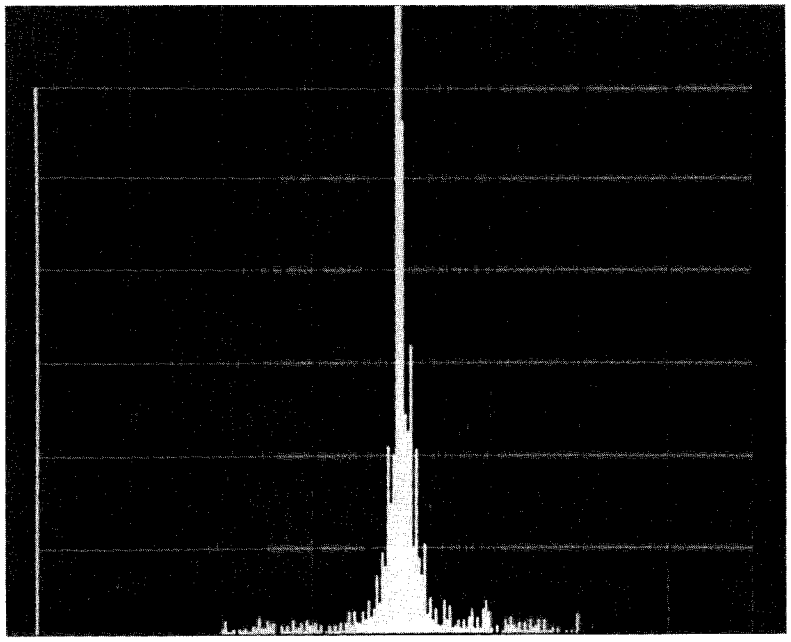

Fig. 12. Histogram of the flow direction of the optical flow vectors in figure 11.

Fig. 13. First frame, miscellaneous hardware sequence.

Fig. 14. Optical flow field obtained from the image sequence of figure 13.

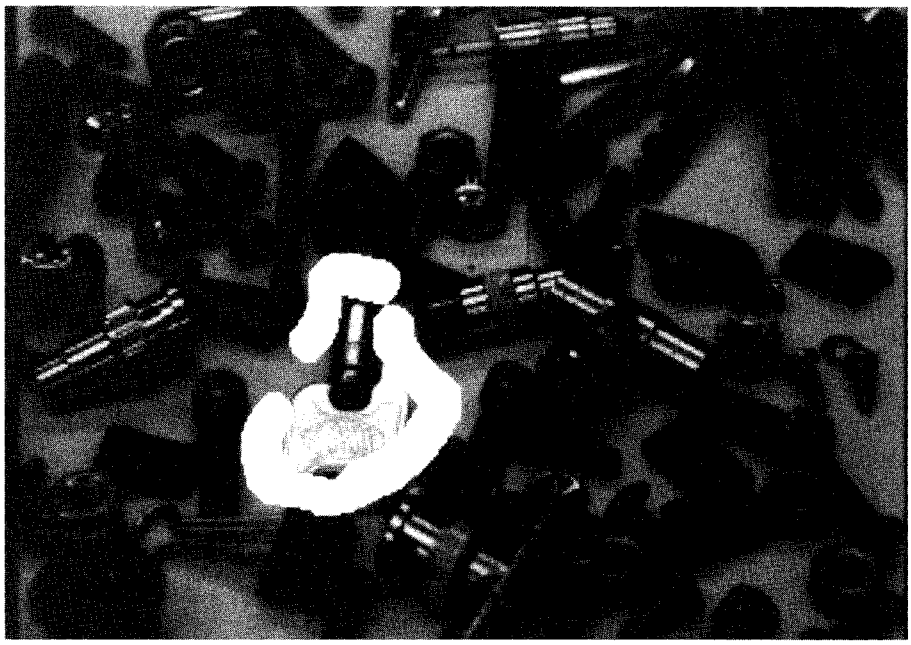

Fig. 15. Boundary of a moving object overlaid onto figure 13.

7.2 Computational Structure

The methods just described can be grouped into three classes. *Point-based* techniques (completely known motion) compare individual optical flow vectors against some standard to determine incompatibilities with the motion of the camera relative to the environment. In all cases described here, the compatibility measure is based on a directional constraint associated with the focus of expansion of the translational flow field. Point-based methods have the advantages of computational simplicity and the ability to detect very small moving objects. They will be most effective when parameters of motion are known precisely and the magnitude of the translational flow field at the point in question is sufficiently large to allow an accurate estimate of direction. *Edge-based* techniques (known rotation, smooth surface) roughly correspond to traditional edge detection. Edge-based motion detection is characterized by the differential flow properties examined and by the filtering technique used to separate edges due to range discontinuities from those due to moving objects. The approach is effective when surfaces are smooth and techniques exist for accurately locating those range discontinuities that do exist. Edge-based methods have the advantage of specifying the outline of moving objects that are detected. They are likely to be of limited use when moving objects are quite small. *Region-based* techniques (tracked object, depth/flow comparisons) examine optical flow values over a region, searching for distributions incompatible with rigid motion. As with edge-based approaches, the viewed region must include portions of both object and environment. As long as the region includes portions of both object and environment, this is an effective test for moving objects and does not require any information about camera motion. The region-based method based on tracking potentially moving objects does not require any information about camera motion, but does require that there be significant variations in range over the visible portions of the environment.

7.3 Limitations

All detection algorithms founded on the motion epipolar constraint share two important shortcomings. First, environmental flow vectors will be small near the FOE regardless of the ranges involved. As a result, detection based on flow orientation will be unreliable within a

Fig. 16. (a) Left image of first stereo pair.

Fig. 16. (b) Right image of first stereo pair.

Fig. 17. Optical flow field obtained for right image sequence of figure 16.

Fig. 18. Disparity field obtained across the stereo pair in figure 16.

Fig. 19. Boundary of a moving object overlaid onto the right image of the stereo pair in figure 16.

region around the FOE.[4] This means that epipolar-based methods will have difficulties for viewing directions close to the direction of motion. This is of course the direction in which moving object detection is likely to be most important. One heuristic for partially overcoming limitations near the FOE is to look for large magnitude values of translational flow near the FOE. Such values correspond either to moving objects or to environmental points that are very close to the camera. Secondly, while the motion epipolar methods were developed to allow for the *possibility* of a moving camera, translational camera motion is actually a *requirement*. Without translational motion, there is no motion epipolar constraint to violate. More specifically, not only must the camera be moving, but significant portions of the visible environment must be sufficiently close to generate detectable nonzero translational flow values. Most methods based on the depth/flow or rigidity constraints should work for both moving and stationary cameras.

No method for detecting moving objects will be effective if it depends on knowing precise values of optical flow. Techniques for estimating optical flow are intrinsically noisy (e.g., see [24]). Additional difficulties arise due to the idealized nature of equations (1)–(3). Real cameras are not point-projection systems. Substantial effort is required to accurately determine the values of x and y in (1)–(3). Geometric distortions in the optical and sensing systems affect measured locations on the image plane. Variabilities in effective focal length can be substantial. Reliable techniques will be based on searching for large-magnitude effects in the flow field [25]. All of the methods described here compare flow vectors to some predetermined standard, or look for significant differences across flow boundaries. As a result, all deal with relatively large-magnitude effects. Reliability is still dependent on scene structure, the nature of camera motion, and position in the visual field relative to the direction of translation. Furthermore, many optical flow estimation techniques perform poorly in the vicinity of discontinuities in flow. Improvements in this regard will lead to more sensitive moving object detection.

[4]Lawton talks about a "dead zone" around the FOE within which no information based exclusively on camera motion is available [23]. This effect is a problem not only for moving object detection, but also for techniques such as motion stereo.

Acknowledgment

This work was supported by AFOSR contract AFOSR-87-0168 and NSF Grants DCR-8500899 and IRI-8722576. Martin Kenner provided significant assistance in preparing the examples.

References

1. G. Adiv, "Inherent ambiguities in recovering 3-d motion and structure from a noisy flow field," *Proc. 3rd IEEE Conf. Comput. Vision Pattern Recog.*, San Francisco, pp. 70–77, 1985.
2. *Proc. Workshop on Motion: Representation and Analysis*, Kiawah Island, SC, May 1986.
3. S. Ullman, *The Interpretation of Visual Motion*, Cambridge, MA: MIT Press, 1979.
4. R. Jain, W.N. Martin, and J.K. Aggarwal, "Extraction of moving object images through change detection," *Proc. 6th Intern. Joint Conf. Artif. Intell.*, Tokyo, pp. 425–428, 1979.
5. R. Jain, D. Militzer, and H.-H. Nagel, "Separating non-stationary from stationary scene components in a sequence of real world TV images," *Proc. 5th Intern. Joint Conf. Artif. Intell.*, Cambridge, MA, pp. 425–428, 1977.
6. A.M. Waxman and J.H. Duncan, "Binocular image flows," *Proc. Workshop on Motion: Representation and Analysis*, Kiawah Island, SC, 1986.
7. W.B. Thompson, K.M. Mutch, and V.A. Berzins, "Dynamic occlusion analysis in optical flow fields," *IEEE Trans. PAMI* 7:374–383, July 1985.
8. R.C. Jain, "Segmentation of frame sequences obtained by a moving observer," *IEEE Trans. PAMI* 6:624–629, September 1984.
9. W.B. Thompson, "Combining motion and contrast for segmentation," *IEEE Trans. PAMI* 2:543–549, November 1980.
10. W.F. Clocksin, "Perception of surface slant and edge labels from optical flow: A computational approach," *Perception* 9:253–269, 1980.
11. K. Nakayama and J.M. Loomis, "Optical velocity patterns, velocity sensitive neurons, and space perception: A hypothesis," *Perception* 3:63–80, 1974.
12. D.J. Heeger and G. Hager, "Egomotion and the sabilized world," *Proc. 2nd Intern. Conf. Comput. Vision*, Tampa, pp. 435–440, 1988.
13. Z. Zhang, O.D. Faugeras, and N. Ayache, "Analysis of a sequence of stereo scenes containing multiple moving objects using rigidity constraints," *Proc. 2nd Intern. Conf. Comput. Vision*, Tampa, pp. 177–186, 1988.
14. A.R. Bruss and B.K.P. Horn, "Passive navigation," *Comput. Vision, Graphics Image Process.* 21(1):3–20, 1983.
15. D.H. Ballard and O.A. Kimball, "Rigid body motion from depth and optical flow," *Comput. Vision, Graphics Image Process.* 22: 95–115, 1983.
16. D.A. Marr, *Vision*, San Francisco: W.H. Freeman, 1982.
17. A. Bandopadhay, B. Chandra, and D.H. Ballard, "Active navigation: Tracking an environmental point considered beneficial," *Proc. Workshop on Motion: Representation and Analysis*, Kiawah Island, SC, pp. 23–29, 1986.
18. T.S. Huang, S.D. Blostein, A. Werkheiser, M. McDonnel, and M. Lew, "Motion detection and estimation from stereo image sequences: Some preliminary experimental results," *Proc. Workshop on Motion: Representatin and Analysis*, Kiawah Island, SC, pp. 45–46, 1986.
19. J.J. Gibson. *The Perception of the Visual World*, Cambridge, MA: Riverside Press, 1950.
20. R.A. Brooks, A.M. Flynn, and T. Marill, "Self calibration of motion and stereo for mobile robots," *Proc. 4th Intern. Symp. Robotics Res.*, 1987.
21. J.H. Reiger and D.T. Lawton, "Sensor motion and relative depth from difference fields of optic flows," *Proc. 8th Intern. Joint Conf. Artif. Intell.*, Karlsruhe, pp. 1027–1031, 1983.
22. S.T. Barnard and W.B. Thompson, "Disparity analysis of images," *IEEE Trans. PAMI* 2:333–340, July 1980.
23. D.T. Lawton, personal communication.
24. J.K. Kearney, W.B. Thompson, and D.L. Boley, "Optical flow estimation: an error analysis of gradient-based methods with local optimization," *IEEE Trans. PAMI* 9:229–244, March 1987.
25. W.B. Thompson and J.K. Kearney, "Inexact vision," *Proc. Workshop on Motion: Representation and Analysis*, Kiawah Island, SC, pp. 15–21, 1986.